Algorithm Development for Column Water Vapor Retrieval Using The SAM Sensor

Joshua Willimas, John DeVore, AT Stair Jr., and Doran Baker

Abstract—To understand and model the energetics of the Sun-Earth connection, measurements of specific atmospheric molecules are beneficial. Our objective is to formulate an algorithm to derive temporally varying atmospheric water vapor concentrations as a function of altitude, latitude, and longitude from solar irradiance absorption measurements. The Visidyne SAM instrument is used to study the size distribution of cloud particles. By introducing a spectrometer to the SAM instrument, column water vapor values are produced as an additional data product. A new model algorithm was developed and tested against existing algorithms. Through a least-squares analysis, the new algorithm showed an improvement of a factor of up to 23 over the industry standard. A test was also conducted to ascertain which water absorption bandpass produces the least error. Through these tests an improved model algorithm has been produced.

I. Introduction

THE Visidyne SAM (Sun and Aureole Measurement) instrumentation system is designed to measure the absorption and scattering of sunlight attributable to H_2O in the Earth's atmosphere. For SAM to measure solar aureoles, the instrument employs a Sun tracker platform that directs the unscattered sunlight into a blackened cavity located in the instrument. This is called the beam dump. In order to measure the absorption of sunlight attributable to H_2O the sunlight in the beam dump was conveyed over to a fiber-optic cable to a spectrometer. This spectrometer provides the data necessary to calculate column water vapor.

Column water vapor (CVW) is a measure of the amount of water vapor in the atmosphere. It represents:

> The total amount of water vapor contained in a vertical column of unit cross section area, expressed as the depth (in cm of precipitable water) of the liquid equivalent of the vapor in the column if all that water vapor were condensed and collected [*Hopkins*, 2008].

Water vapor absorbs in a number of wavebands of light in the visible and near-infrared spectral region. This paper addresses the development of an improved model algorithm to accurately ascertain column water vapor from the absorption spectra. To accomplish this development the following requirements needed to be addressed:

1) Research existing methods and techniques of column water retrieval.
2) Development of an improved algorithm for the SAM instrument.
3) Validation of the measurements by minimum mean squared error tests.

4) Formulate, test, and document the improved model algorithm.

Through these steps, using experimental data, a new algorithm was tested against existing algorithms, and these tests demonstrated the magnitude of improvement obtained through using the new approach. Upon accomplishing this, the advanced algorithm was tested to discover under what conditions it produces the most accurate results. Both of these tests were conducted using a least-squares approach. By minimizing the error a model algorithm under appropriate constraints was formulated. This development provided a reliable data set to ascertain the distribution and dynamics of column water vapor, and in turn further the understanding of climate change trends.

II. Algorithm Development

Studies to develop more accurate methods to measure atmospheric column water vapor have been conducted over the last few decades [*Reagan et al.*, 1995]. Innovative approaches have significantly improved the validity of these measurements. These improvements have come both in the form of technology and as improved algorithms used to ascertain the column water vapor. Measurement of the irradiance of the light outside the Earth's atmosphere in comparison with the irradiance at the Earth's surface leads to the determination of the column water vapor in the atmosphere. This absorption is the basis of the column water retrieval method.

Research to develop an optical technique to measure atmospheric column water vapor has been studied at the University of Arizona [*Thome et al.*, 1993]. Algorithms were developed using two separate wavelength bandpasses centered at 820 and 940 nm. The wavelength bandpass was chosen to cover a specific water absorption band. The center of the second bandpass was positioned close to, but not overlapping the water band. This channel is known as the guard band. The derived algorithm, [*Thome et al.*, 1993], uses a ratio of the responses in these two bands in the calculation of the concentration of column water vapor in the atmosphere.

The following symbols will be used throughout the derivation of the column water vapor algorithm:

CWV	Column water vapor
W	Center wavelength of water band
G	Center wavelength of guard band
$R(\lambda)$	Irradiance measured
$R(W)$	Irradiance at water band
$R(G)$	Irradiance at guard band
$R_O(\lambda)$	Solar spectral irradiance outside atmosphere
$R_O(W)$	Irradiance outside atmosphere at water band
$R_O(G)$	Irradiance outside atmosphere at guard band
O_R	Irradiance Ratio $\frac{R_O(W)}{R_O(G)}$
$\tau_a(\lambda)$	Spectral optical depth for aerosol particles
$\tau_R(\lambda)$	Spectral optical depth for Rayleigh scattering
$\tau_3(\lambda)$	Spectral optical depth for O_3
τ_R	Rayleigh scattering $\tau_R(G) - \tau_R(W)$
d	Earth to Sun distance (in astronomical units)
u	Column water-vapor (cm)
m	Relative air-mass
a	Multiplicative parameter
b	Exponential parameter
c	Additive parameter

At least two different algorithms are in use to retrieve CWV. These are derived by using Beer's law [*Tissue*, 2008] in conjunction with an empirical model of water vapor absorption. The two algorithms use two different empirical models to represent the optical absorption by water vapor at specific wavelengths. These empirical models are shown in equations 1 and 2.

$$\ln(R(\lambda)) + m(\tau_a(\lambda) + \tau_R(\lambda) + \tau_3(\lambda)) = \\ \ln(R_0(\lambda)d^{-2}) + a(mu)^b. \quad (1)$$

$$\ln(R(\lambda)) + m(\tau_a(\lambda) + \tau_R(\lambda) + \tau_3(\lambda)) = \\ \ln(R_O(\lambda) * d^{-2}) + c + (mu)^b. \quad (2)$$

Each of these two equations represents a different method to model water vapor absorption using Beer's law [*Thome et al.*, 1993]. The derivation approach for each of these models are the same, therefore only the multiplicative model derivation will be given.

The representations defined by equations 1 and 2 must be compared with a guard band. A guard band is chosen near an appropriate spectral water band feature. This is done to minimize responsivity and instrument noise differences between the two optical measurement channels. A small wavelength separation also minimizes any differences in the sunlight extinction in the two spectral bandpasses other than for water vapor. The guard band is selected such that the relative difference between the two bandpasses is a measure of the water vapor absorption. Equation 3 gives the representation of the guard band.

$$\ln(R(\lambda)) + m(\tau_a(\lambda) + \tau_R(\lambda) + \tau_3(\lambda)) = \\ \ln(R_O(\lambda)d^{-2}). \quad (3)$$

Comparing equations 1 and 3 demonstrates the difference between the water and guard bands. With only a small wavelength separation, parameters $\tau_a(\lambda)$, $\tau_R(\lambda)$, and $\tau_3(\lambda)$ are reasonably comparable for the water and guard bands. This

observation is the basis of the algorithm. In order to solve for the parameters a, b or b and c, which represent the water vapor model, namely equations 2 and 3 which are combined to form equation 4.

$$\ln(\tfrac{R(G)}{R(W)}) + m(\tau_a(G) + \tau_R(G) + \tau_3(G) - \\ \tau_a(W) - \tau_R(W) - \tau_3(W)) = \ln(\tfrac{R_O(G)}{R_O(W)}) + a(m*u)^b. \quad (4)$$

The center wavelength selection for the guard band in proximity to that of the water band facilitates some key assumptions which can be made. When emissions at two similar wavelengths are compared, the different reflections and refractions due to $\tau_a(\lambda)$, $\tau_R(\lambda)$ and $\tau_3(\lambda)$ are comparable in values. This assumption is implicit in the simplification of equation 4 to yield equation 5.

$$\ln(\frac{R(G)}{R(W)}) = \ln(\frac{R_O(G)}{R_O(W)}) + a(m*u)^b \quad (5)$$

The result of solving for the column water vapor u given equation 6.

$$u = \frac{1}{m}(\frac{\ln(O_R * R(G)/R(W))}{a})^{\frac{1}{b}}. \quad (6)$$

This equation is the representation for the multiplicative model given in equation 1. By a similar derivation, equation 7 derives from the additive model given in equation 2.

$$u = \frac{1}{m}(\ln(O_R * R(G)/R(W)) - c)^{\frac{1}{b}}. \quad (7)$$

III. NEW THREE-PARAMETER ALGORITHM

The approach to finding an improved model and method of approximating CWV is again to use Beer's law. By employing an improved water vapor absorption model, the error will be minimized between the theoretical and the measured values. Equations 1 and 2 are two different representations of water absorption used in combination with Beer's law. Upon a review of these two equations, tradeoffs were found in approximating the water vapor absorption. This led to the development of a new three-parameter approach. With the combination of the two dual-parameter systems, a third degree of freedom is introduced into the system. This third parameter enables the algorithm to better approximate the water absorption. This third parameter adds some complexity to the overall system, but each parameter set has to be calculated only once for each choice of spectral bandpasses. See section V for comparison of the two- versus three-parameter approaches.

The derivation of the three-parameter approach is similar to that for each of the two-parameter approaches. Differences with preceding algorithms include:

1) A square-root dependance for the parameter b is not presumed.
2) It is not assumed that the Rayleigh scattering in the water and guard bands are the same. In some previous models it is assumed that the Rayleigh scattering is the same for the water and guard bands [*Thome et al.*, 1993].
3) All three parameters a, b, and c are included in approximating water absorption.

4) Equation 8 will be used to represent the water band, and equation 3 is used to represent the guard band.

$$\ln(R(\lambda)) + m(\tau_a(\lambda)) + \tau_R(\lambda) + \tau_3(\lambda) =$$
$$\ln(R_O(\lambda) * d^{-2}) + c + a(mu)^b. \qquad (8)$$

The Rayleigh scattering is similar in the water and guard bands, but due to different wavelength locations of these bands the Rayleigh scattering values can be slightly different. For this step in the derivation, as a simplification it is assumed, the optical scattering due to O_3 and aerosols, namely $\tau_a(\lambda)$ and $\tau_3(\lambda)$, are the same for both bands. In future work, if these scattering effects are measured and not assumed to be approximately the same, a more accurate measure of CWV will result. This will be pursued in future versions of the algorithm. Using these assumptions, equation 9 results from the combination of equations 3 and 8.

$$\ln(\tfrac{R(G)}{R(W)}) + m(\tau_R(G) - \tau_R(W)) =$$
$$\ln(\tfrac{R_O(G)}{R_O(W)}) + c + a(m * u)^b. \qquad (9)$$

Through simplification, equation 9 becomes equation 10.

$$\ln(\tfrac{O_R * R(G)}{R(W)}) + m(\tau_R(G) - \tau_R(W)) =$$
$$\ln(\tfrac{R_O(G)}{R_O(W)}) + c + a(m * u)^b. \qquad (10)$$

Equations 9 and 10, produced by instituting the new three-parameter approach, are more complex than the current two-parameter approach. However, through extensive testing an improved solution was formulated. Equation 11 gives the resulting CWV equation model for this approach.

$$u = \frac{1}{m}\left(\frac{\ln(O_R * R(G)/R(W)) + m * \tau_R - c}{a}\right)^{\frac{1}{b}}. \qquad (11)$$

IV. ALGORITHM VALIDATION AND COMPARISON

After conceptual development of the algorithm was accomplished the next steps were implementation and testing. For both the two- and three-parameter approaches (equations 6, 7, and 11) a method was formulated to solve for each of the parameters (a, b, and c). Once formulated, a comparison was conducted to ascertain which is the best algorithm. These tests showed under what circumstances which algorithm performs the best, *i.e.*, what wavelengths and spectral bandwidths are best to use.

The validity of each CWV algorithm is, in part, determined by how the parameters a, b, and c are formulated to model the water vapor absorption. The process [*Schmidis et al.*, 2001] sensitive to parameter values, and these parameters exhibit a large CWV variation with even slight alterations in each method of retrieval. Consequently, a method needed to be formulated to calculate these constants for each different retrieval technique. Parameters, including optimal bandwidth, center wavelength, and characterization of the measurement instruments were taken into account.

Independent of the sensor used to measure irradiance in the water and guard bands, a method needed to be found to establish the values of a, b, and c which result in the smallest error. This error is defined as the correct value, produced by the atmospheric model, minus the calculated value, produced by the algorithm. The error value determines which parameters best fit the actual values of CWV. Under realistic conditions a nonzero error is expected. The objective of the present analysis is to minimize this error for the existing model. To find this minimum, a least-squares error analysis was utilized.

In the formulation of the CWV algorithm the three parameters a, b, and c were used. A least-squares analysis could be conducted for each of the three parameters, but *a priori* knowledge simplifies the process and helps constrain the results to stay within physical bounds. When conducting a least-squares analysis a standard training data set is used in the calculation. Since an infinite set of data is not possible, the least-squares analysis merely approximates the parameters which will result in the smallest error. However, the parameter values that determine this error can lie outside the set of real possibilities.

In this study fewer than a hundred points of training data were used. Such a limited set of data points could cause the parameter values, which result in the smallest error, to lie outside of their acceptable ranges. From the understanding of the optical model used to approximate the water vapor absorption, it is known that the exponent b must physically lie between zero and two. This *a priori* knowledge helps ensure that each calculated parameter is constrained to the appropriate range.

To initiate the least-squares analysis, the error value is squared at each data point as defined by equation 12.

$$Error^2 = [c + a(m * u)^b - (ln(\tfrac{O_R * R(G)}{R(W)}) + m * \tau_R)]^2. \qquad (12)$$

This is a representation of the error squared at a given instant of time. The squared error was used in this analysis because minimizing the error is the equivalent of minimizing the error squared. Accordingly, the mean squared error is subsequently referred to as 'the error'. To calculate the mean squared error (MSE), equation 12 is summed over the number of measurements, and divided by the number of measurements. The MSE is represented by equation 13.

$$MSE =$$
$$\tfrac{1}{n}\sum_{i=1}^{n}[c + a(m_i * u_i)^b - (ln(\tfrac{O_R * R(G)_i}{R(W)_i}) + m_i * \tau_R)]^2 \qquad (13)$$

The MSE shown in equation 13 is the difference between the value of the actual water vapor absorption and the estimated value. It is noted that this is a different value than the MSE of the CWV. By minimizing the absorption MSE, in turn, the CWV MSE is minimized. These values will be different, but represent the same error. The magnitude of MSE introduced by the estimated absorption is proportional to the CWV MSE. For the derivation of the minimum mean squared error (MMSE), the absorption MSE is used as shown in equation 13. After the

derivation the two are distinguished by the titles 'absorption' and 'CWV MSE'.

To minimize the MSE given by equation 13, derivatives are taken with respect to parameters a and c. As stated earlier, the range of b is defined to lie between zero and two. For this part of the analysis, it is assumed that the value of b is known. The solution for the correct value of b is given later in the analysis. Taking the derivatives of the MSE with respect to a and c yields equations 14 and 15.

$$\frac{\partial MSE}{\partial a} = \frac{2}{n} \sum_{i=1}^{n}([c + a(m_i * u_i)^b - (ln(\frac{O_R * R_i(G)}{R_i(W)}) + m_i * \tau_R)]^2 * (m_i * u_i)^b). \quad (14)$$

$$\frac{\partial MSE}{\partial c} = \frac{2}{n} \sum_{i=1}^{n} [c + a(m_i * u_i)^b - (ln(\frac{O_R * R_i(G)}{R_i(W)}) + m_i * \tau_R)]^2. \quad (15)$$

To determine the MMSE with respect to a and c, both equations 14 and 15 are set to zero. This results in two equations and two unknowns. Equation 16 shows a simplification of equation 14, and equation 17 shows the simplification of equation 15.

$$c \sum_{i=1}^{n}(m_i * u_i)^b + a \sum_{i=1}^{n}(m_i * u_i)^{2b} = \sum_{i=1}^{n}([ln(\frac{O_R * R_i(G)}{R_i(W)}) + m_i * \tau_R] * (m_i * u_i)^b). \quad (16)$$

$$c \sum_{i=1}^{n} 1 + a \sum_{i=1}^{n}(m_i * u_i)^b = \sum_{i=1}^{n}[ln(\frac{O_R * R_i(G)}{R_i(W)}) + m_i * \tau_R] \quad (17)$$

At this point it is possible to solve for both a and c. In order to solve this set of equations, they are placed into matrix form as given in equation 18.

$$\begin{bmatrix} \sum_{i=1}^{n}(m_i * u_i)^b & \sum_{i=1}^{n}(m_i * u_i)^{2b} \\ \sum_{i=1}^{n} 1 & \sum_{i=1}^{n}(m_i * u_i)^b \end{bmatrix} * \begin{bmatrix} c \\ a \end{bmatrix} = \begin{bmatrix} \sum_{i=1}^{n}[ln(\frac{O_R * R_i(G)}{R_i(W)}) + m_i * \tau_R](m_i * u_i)^b \\ \sum_{i=1}^{n}[ln(\frac{O_R * R_i(G)}{R_i(W)}) + m_i * \tau_R] \end{bmatrix}. \quad (18)$$

Having obtained a least-squares solution for both a and c, we can return to the earlier assumption of knowing the value of b. Understanding the nature of the exponential parameter in this function demonstrates acceptable ranges of values for b. The range in this case is limited to between zero and two. To find the best method of solving for this parameter, a training data set is needed which contains the following:

1) Water band measured value.
2) Guard band measured value.
3) Corresponding CWV value.

Using this training data set, a comparison between the actual CWV and the calculated value can be produced. The method used in this paper is the MMSE. This is accomplished by setting b to a fixed value, solving for both a and c, and then calculating the MSE for all values of u given by the training data set. This was done for incremental values of b between zero and two. The MMSE's are taken as the values of b which best fit the data. This method also produces the corresponding values for a and c. This process is shown in equation 19.

$$MMSE = \min_b \sum_{i=1}^{n}(u_i(calc) - u_i(actual))^2, \\ 0 < b <= 2. \quad (19)$$

All derivations shown in this section have been formulated for the three-parameter approach. This same method is used to solve for the each of the two-parameter approaches. However, there is a slight modification in which there is only one parameter for which the least squares analysis is conducted.

V. Two- Versus Three-Parameter Approach

Each of the two- and three-parameter approaches have been derived in detail. These derivations provide the means by which a comparison can be made to ascertain which provides the most accurate representation of water vapor absorption. This comparison was conducted over multiple center wavelengths and multiple bandwidths. It was done to show the validity of the results and to assess the magnitude of improvement of one approach over the other. Differences between the different two-parameter models were considered, but it was found that the multiplicative model produced the least MSE, and therefore this model was used in this analysis [*Schmidis et al.*, 2001]. For this reason, only the multiplicative version of the two-parameter approach will be compared with the three-parameter approach herein.

Table I demonstrates the magnitude of improvement for each of the considered water vapor bands. This led to the following determinations with respect to the two- versus three-parameter approaches:

1) For each of the water-bands studied, the performance of the three-parameter approach is equal to or exceeds that of the two-parameter.
2) Both model approaches produce the smallest MMSE's at center water absorption bands of 820 and 940 nm.
3) At these center wavelengths with a bandwidth less than 10 nm the three-parameter model outperforms the two-parameter model by at least a factor of 10.
4) The three-parameter approach performs best over the wide range of possible values of CWV.
5) At 940 nm with a bandpass of 10 nm, namely AERONET's bandpass, the three-parameter approach demonstrated an improvement by a factor of 22.

VI. Conclusions

The following are some of the conclusions which were formulated from the research reported in this paper:

1) From the forgoing observations, the conclusion can be drawn that, for all wavelengths, the three-parameter water vapor absorption model is the best choice.

2) Research into already-available algorithms gave improved insight into approximating water vapor absorption. The combination of two separate water vapor absorption models demonstrated improvements by a factor of 22 over the available models' individual performance using a three parameter fitting approach. Through a minimum mean squared error test, the improvement of an improved algorithm over the original model was demonstrated for each of the 720, 820, 940, and 1465 nm water vapor absorption bands. Consequently, using the specifications of the national AERONET network sensor [*Halthore et al.*, 1997] at 940 nm, the factor of improvement of 22 was validated. Such a significant improvement demonstrated that the three-parameter algorithm model was a significant improvement.

3) Minimum mean squared error tests were conducted for the 720, 820, 940, and 1465 nm absorption bands. These results were tabulated to select the best water band to employ in the SAM instrument. The results also demonstrated for each of the water vapor absorption bands, what corresponding spectral bandwidth would give the best results.

4) The improved model algorithm was tested and documented.

5) Characterization parameters a, b, and c were calculated and tabulated.

ACKNOWLEDGMENT

REFERENCES

[1] Martin Chaplin, "Water Absorption Spectrum," [http://www.lsbu.ac.uk/water/vibrat.html], Jan 2008.

[2] G. Herzberg, *Molecular Specta and Molecular Structure*. Litton Educational Publishing, 1945, vol. II Infrared and Raman Spectra of Polyatomic Molecules.

[3] Edward J. Hopkins, "RADIOSONDE OBSERVATION (RAOB) DATA TABULATION," [http://www.aos.wisc.edu/ hopkins/aos100/raobdoc.htm], Feb 2008.

[4] N. Robinson, *Solar Radiation*. Elsevier Publishing Company, 1966.

[5] J. Reagan, "A comparison of columnar water vapor retrievals obtained with near-ir solar radiometer and microwave radiometer measurements," *American Meteorological Society*, vol. 34, pp. 1384–1391, Jun 1995.

[6] K. J. Thome, "Determination of precipitable water from solar transmission," *American Meteorological Society*, vol. 31, pp. 157–165, Feb 1993.

[7] Brian M. Tissue, "Beer-Lambert Law," [http://elchem.kaist.ac.kr/vt/chem-ed/spec/beerslaw.htm], Mar 2008.

[8] NASA Goddard Space Center, "AERONET (AErosol RObotic NETwork) program," [http://aeronet.gsfc.nasa.gov/], Jan 2008.

[9] B. Schmid, "Comparison of columnar water-vapor measurements from solar transmittance methods," *Applied Optics*, vol. 40, no. 12, pp. 1886–1896, April 2001.

[10] R. N. Halthore, "Sun photometric measurements of atmospheric water vapor column abundance in the 940-nm band," *Journal of Geophysical Research*, vol. 102, no. D4, pp. 4343–4352, Feb 1997.

[11] U.S. Air Force Research Lab, "Fact Sheet, Modtran 4 Software," [http://www.kirtland.af.mil/library/factsheets /factsheet.asp?id=7915], Jan 2008.
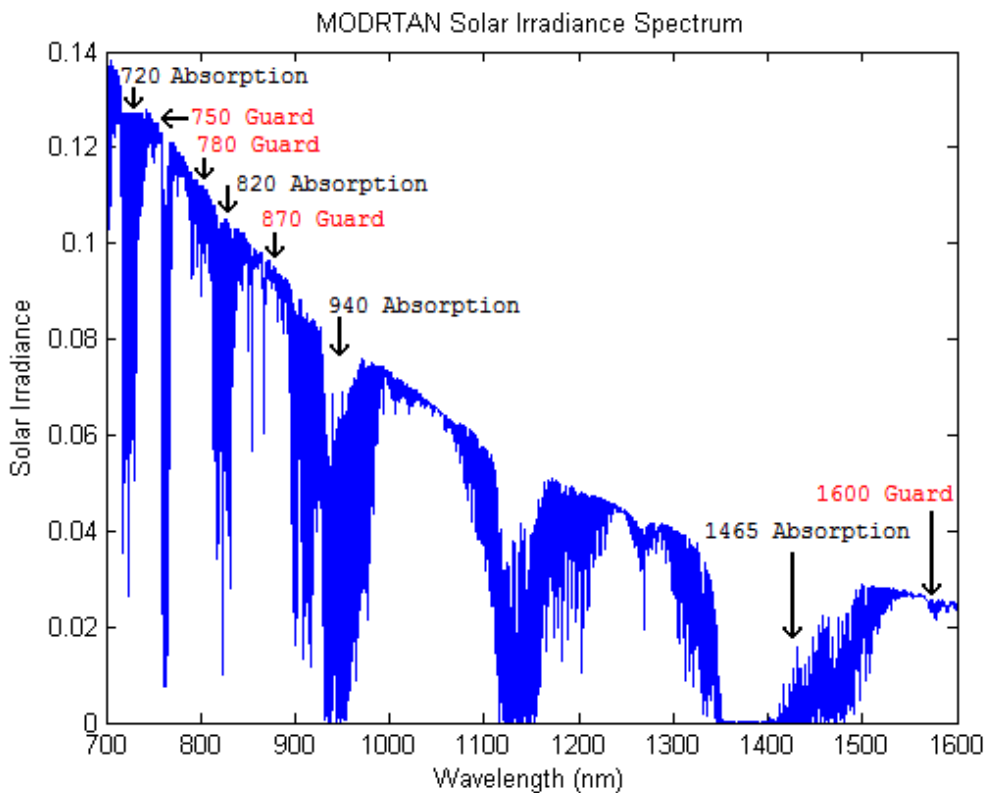
Fig. 1. Water absorption bands.

TABLE I
TWO- VERSUS THREE-PARAMETER COMPARISON WITH THE MAGNITUDE OF CHANGE CORRESPONDING WATER BANDS.

Water (nm)	Guard (nm)	Bandwidth (nm)	Two- MMSE	Three- MMSE	Improvement Factor
1465	1600	1	0.0784	0.0313	2.5
1465	1600	5	0.0158	0.0157	1.0
1465	1600	10	0.0258	0.0149	1.7
940	870	1	0.0939	0.0033	28.5
940	870	5	0.0473	0.0021	22.5
940	870	10	0.0470	0.0021	22.4
820	780	1	0.0180	0.0017	10.6
820	780	5	0.2171	0.0010	217.1
820	780	10	0.1158	0.0134	8.6
720	750	1	0.0535	0.0175	3.1
720	750	5	0.2930	0.0227	12.9
720	750	10	0.3698	0.0253	15.7